MULTIMEDIA INTEGRATED MODELING FOR ENVIRONMENTAL PROTECTION: INTRODUCTION TO A COLLABORATIVE FRAMEWORK

JOHN M. JOHNSTON[1], JOAN H. NOVAK[2*] and STEPHEN R. KRAEMER[1]
*[1]Ecosystems Research Division, 960 College Station Rd., Athens, GA 30605 USA; [2]Atmospheric Modeling Division, MD-80, RTP, NC, 27711 USA, U.S. Environmental Protection Agency (EPA) National Exposure Research Laboratory; * On assignment from the National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce*

Abstract. The EPA's Office of Research and Development is embarking on a long term project to develop a Multimedia Integrated Modeling System (MIMS). The system will have capabilities to represent the transport and fate of nutrients and chemical stressors over multiple spatial and temporal scales. MIMS will be designed to improve the environmental management community's ability to evaluate the impact of air and water quality and watershed management practices on stream and estuarine conditions. The system will provide a computer-based problem-solving environment for testing understanding of multimedia (atmosphere, land, water) environmental problems, such as the movement of chemicals through the hydrologic cycle, and the response of aquatic ecological systems to land-use change, with initial emphasis on the fish health endpoint. The design will attempt to combine the state-of-the-art in computer science, system design, and numerical analysis (i.e., object-oriented design, parallel processing, advanced numerical libraries including analytic elements) with the latest advancements in process level science (hydrology, atmospheric sciences, chemistry, ecology). The purpose of this paper is to introduce a vision for a MIMS and anticipate the challenges to its development.

1. Introduction

1.1 MULTIMEDIA SCENARIO/FUTURE USE CASE

An EPA staff scientist in Region IV wants to prioritize the allocation of funds and effort to solve an environmental problem. The year is 2008 and once again there are water quality concerns in the Albemarle-Pamlico basin on the mid-Atlantic coastline. Fish kills are occurring in the Neuse River estuary in the late summer. Populations of striped bass are quite low in the Albemarle and Pamlico Sounds. Sampling by the U.S. EPA and U.S. Geological Survey (USGS) has detected the presence of heavy metals, including mercury, and pesticides in both Sounds, as well as elevated suspended sediment levels. The presence of the dinoflagellate *Pfiesteria* has also been confirmed. It is known from studies of patterns of human commercial and suburban development, and their impacts on adjacent waterbodies, that nutrients and sediments are also possible stresses on aquatic communities, including striped bass populations. The regional scientist wants to know several things: 1) What factor or combination of factors are contributing to the fish kills in the Neuse estuary, and can these factors be prioritized as they relate to the

Environmental Monitoring and Assessment **63:** 253–263, 2000.

health of the living resources in the Albemarle-Pamlico Sounds?; 2) What are the expected fish population sizes in the Albemarle-Pamlico sounds, given the status of available habitat and the levels of human disturbance?; 3) How are human activities in the basin and airshed related to the ecological response in the estuary, and what control measures or management strategies can be used to improve the situation?; 4) What will these strategies cost? and 5) What chance of success is associated with each?

Non-point source pollution is believed to be the primary stressor in this situation since the majority of point source dischargers are regulated. The pollutants of concern include fertilizers, pesticides, and sediments. Previous scenarios that attributed fish population decline to a single stressor were determined to be inadequate in their analysis of causation. Namely, fish species normally able to survive in the presence of a parasite or pathogen such as *Pfiesteria* were more susceptible when other stressors were co-occurring.

The EPA scientist, sitting at her personal computer in Atlanta, GA, uses the Internet to reach the EPA Multimedia Integrated Modeling System. The MIMS, released for public use in 2008, is a problem-solving environment that enables the client to build and apply modeling and analysis tools specific to this region and scenario. MIMS is a software architecture that provides the tools necessary to both understand complex multi-stressor ecosystem impacts and to assess the cost effectiveness of proposed management strategies. MIMS servers are located in Athens, GA, Las Vegas, NV, Cincinnati, OH, and Research Triangle Park (RTP), NC, and are utilized depending on the degree of computational usage and network traffic. From this client-server connection the client selects the necessary models for this problem: coupled atmospheric-hydrospheric models, ecological models for the aquatic communities and bioaccumulation, and a land-use model that can simulate changes in land cover and the resulting effects on watershed hydrology and water quality. The data for parameterizing these models are also located and linked. These include hydrodynamic and water quality data sufficient to parameterize the distributed hydrologic models, land cover data, and 20 years of striped bass population data. The U.S. Fish and Wildlife Service home page reveals that several Internet web sites provide fish community and other aquatic ecology data for the region via immediate file transfer. Researchers at the University of North Carolina Marine Laboratory have been collecting long term data for river characterization (turbidity, salinity, and material transport in and out of the channels) that will provide a baseline against which to perform habitat comparisons. Upon checking the Internet for data dealing with the effects of current and historic development patterns along the riverine and estuarine coastlines, she discovers that the National Marine Fisheries Group (Beaufort, NC) has coverages available for download that detail coastal development patterns and the accompanying hydrologic and water quality data.

From the analysis, it is determined that the striped bass populations are, in fact, expected to be at higher levels than are currently observed (and the long term data exist to corroborate this). When attempting to evaluate the causative factors, it is not possible to separate the effects of sediment runoff and pesticide loading to the receiving waters in each sound. Forested riparian buffer strips are found to be effective in reducing pesticide and sediment loads significantly, and an assessment of nitrogen pollution reveals that this problem can only be addressed effectively by including non-point-source management options. In addition, through bass population and aquatic community modeling, it is shown that the creation of well-designed riparian habitats also increases striped bass population recruitment. The next steps in this scenario would involve using the modeling framework to test riparian buffer zones of various sizes in different locations to estimate the cost and the incremental degree of improvement in specific reaches of higher order streams and river systems.

2. A brief history of multimedia modeling within U.S. EPA. Where are we coming from?

A synopsis of the modeling practices of the U.S. EPA provides an important perspective on the current status of multimedia modeling within the U.S. EPA Office of Research and Development (ORD). It should be mentioned that our use of the term "U.S. EPA model" denotes research models developed by ORD scientists and their collaborators. With the exception of models used by the Office of Air that have gone through the listing process in the Federal Register, such models are not "officially approved" and can be considered to be research tools in one form or another. We acknowledge that there are many models that are routinely and extensively used by U.S. EPA national programs, regions, other departments, agencies and consultants. However, there is currently no list of EPA approved models, though there is interest in standards for model development and testing under the Models 2000 Initiative.

The way that U.S. EPA was organized from the beginning, in response to enabling legislation from Congress, has had an effect on both modeling science and software development. Program offices were established for distinct media: e.g., the Office of Air, Office of Water, and the Office of Solid Waste. For the most part this structure (though nicely corresponding to Federal law) did not promote modeling that addressed the need to track mass and energy fluxes and concentrations of pollutants in and through each of the media for a holistic assessment. In this setting the ecosystem was not envisioned as a whole but rather as separate components for protection that might be eventually integrated. This single medium emphasis also had the effect of removing pollutants from one medium by merely transporting them to another. Examples of this include solid wastes such as sludge from wastewater treatment plants, toxic ash from control devices on air emissions (Theodore and Theodore 1996), and agricultural best management practices that reduce surface water runoff by transfer of pollutants to subsurface recharge.

Given this organizational structure, there was no common conceptual model for researchers to use in order to describe the complete system and specify the types of modules to be created. The research emphasis in the past has been on models that predicted the fate and transport of toxic compounds. Because the chemistry and physics of pollutant speciation, transformation, and transport had a relative wealth of laboratory and field data from which to draw, modeling activities in these areas were more intensive. The system-wide cycling of toxic compounds and other materials was also seemingly ignored. Procedural languages such as FORTRAN were also the most widely used for software development, so code maintenance and adaptations of larger models was an intensive and expensive enterprise. Legacy codes have the benefits of long term testing and stability, name recognition and confidence, but they can lock-in perhaps out-dated science, and might be difficult to adapt to an evolving ecological conceptual model.

The first generation of air pollutant models were specifically transport and decay models, with no accounting for interactions between chemicals. The dispersion of a single species such as SO_2 from a stack would be simulated by a gaussian (distribution) plume. For the most part, the species were all conservative with a simple decay rate factored in. The successive generations of air models were gridded spatially and included transformation reactions (such as photochemistry) as well as transport. However, these models were scale and pollutant specific. Because of the desire to include more realistic chemistry and physics, as well as to demonstrate the spatial aspects of each system, this generation of models was quite difficult to build, manage, and maintain. Much was learned from this experience that created many different tools to solve a variety of environmental management questions.

3. Present multimedia modeling emphasis: Where are we now?

With the third and present generation air models, all of the capabilities of the individual models were implemented into a single system by taking a "one atmosphere" approach to the science formulation, and by using object-oriented analysis/design for a flexible software framework to deal with the complexities of an integrated system. Models-3 (USEPA 1998) is a problem-solving environment designed to simplify the use and continued evolution of environmental models. Models-3 contains components that build, apply, and display results from air quality models. The initial implementation of Models-3 contains a Community Multi-scale Air Quality (CMAQ) modeling system for urban to regional scale air quality simulation of tropospheric ozone, acid deposition, visibility, and fine particles (USEPA 1999). Models-3 and CMAQ, in combination, form a powerful tool that enables the client to specify key attributes of the modeling domain of interest, such as spatial and temporal extent and resolution, coordinate system definition, and map projection. Models-3/CMAQ also allows the client to select from a collection of science process modules to build customized models specific to their problem or the computing resources available. Within the U.S. EPA this system represents the state-of-the-art in software engineering and provides a foundation for continual evolution of complex environmental modeling systems.

As it currently exists, Models-3 is a potential multimedia framework that has been implemented only for air. The next logical step is the integration of selected hydrology, and aquatic and terrestrial ecosystem models into Models-3 as the starting point for development of an environmental problem-solving environment for performing ecological risk assessments. However, review of several popular water quality models identified the following issues. Some models are limited to a specific geographic area or problem (such as point models) and require extensive data collection and calibration to ensure utility and applicability to the area of interest. This heavy reliance on calibration in a specific area limit the models' use for predictive simulation and effective linkage to air quality or ground water models. There needs to be a shift to more physically-based models to enable more realistic linkages that deal with the time and space scale differences among the media. Many of the current models oversimplify the dimensionality of the chemical and physical exchanges leading to one-dimensional analysis and statistical Monte Carlo approaches instead of full three-dimensional simulation. Data and computational power limitations can be credited as factors that led to these limitations in model design.

Additional limitations arise beyond weaknesses in model formulation when linkage and data exchange between atmospheric, aquatic, and terrestrial models are considered. Modeling studies and observations in the Chesapeake Bay have shown the multimedia nature of nitrogen loadings; airborne sources of nitrogen to the watershed may account for 25–30% of the nitrogen load to the Bay (U.S. EPA Chesapeake Bay Program); and ground water studies indicate that base-flow nitrate load accounted for 26 to about 100% of total stream flow nitrate load, with a median value of 56%, and it accounted for 17 to 80% of total stream flow total-nitrogen load, with a median value of 48% (Bachman et al. 1998). However, it is difficult to link models across, and even within media. For example, many water quality models have overlapping functionality and different assumptions embedded within the code. Semantic differences across disciplines create additional confusion. In addition, modelers focused on one media tend to simplify the inputs from other media in a cross-media model linkage due to inexperience with the details of the other disciplines. Is it realistic to expect significant progress in more holistic cross-media model development if each modeler must become somewhat of an expert in the other scientific disciplines being linked?

Our awareness of environmental problems is growing in complexity and scope. Local management solutions alone are no longer able to address many of today's problems. Deeper impacts are being made with greater persistence than ever before. The economy is intricately linked to proposed environmental management solutions, thus many of these problems cannot be solved and instead must be continuously managed to balance human and ecosystem health with the economics of a region. Accordingly, the models used to assess these problems and evaluate alternative solutions are increasing in complexity, yet they must remain accessible and easy to use by a growing number of organizations. The time has come to consider collaborative rather than competitive approaches to significantly advance the state-of-the art. The time is right—computing advances and remote sensing techniques have progressed to the point where it may be feasible to collect and process the data necessary to support more physically- based first-principles simulations.

4. Multi-media Integrated Modeling System: How do we get there?

U.S. EPA is embarking on a long term project to foster a more holistic approach toward environmental modeling by promoting collaboration, organization, communication, and evolution toward a multi-discipline problem-solving environment. There are two parallel research and development efforts towards the goal of creating a MIMS. First, scientific emphasis will be on physically-based representation of our understanding of the movement of solutes and particles at multiple time and space scales and media (air-land-water). Primary transport and transformations will include representation of the hydrologic cycle, biogeochemical cycles, and the resultant advective flow, accounting for the water budget and mass conservation of the pollutants and nutrient budgets. Incorporating these advances in multimedia modeling, changes in environmental state can be used in combination with ecological process and effects models leading to estimation of ecological risk based upon exposure of ecological receptor(s) to stressor(s). With the realization that ecosystems and ecological problems are more complex than any single path model or single stressor can demonstrate, the emphasis has changed to model development which attempts to embrace the complexity: multiple ecological endpoints, multiple pathways, and multiple temporal and spatial scales.

The second research focus involves the development of a MIMS within the context of performing ecological risk assessment and management of aquatic systems at the watershed scale. In essence, there are two distinct and compatible areas of research within this program. MIMS is obviously needed as a research tool for those who hope to improve the state-of-the-art in fully geographically-distributed multimedia physical and chemical modeling. However, such science, as well as its technological expression of uncertainties, is needed in an applied context for ecological risk assessment. Done properly, a MIMS can thereby meet both the needs of the research and management communities.

Development of a conceptual process model detailing component interactions is ongoing, beginning with a description of the hydrologic and nitrogen cycles as the starting point for object-oriented analysis and design. Integral to the vision of interdisciplinary cooperation on multi-media modeling is the participation of remote, multiple clients working on the conceptual model. Communication between the various individuals and organizations is essential and necessitates an electronic medium, such as Lotus Notes/Domino or electronic news groups for discussion of issues. Software development with an eye to reuse is also a priority. Not only does a modular approach to software design improve efficiency, it also makes more use of the research that is available at any given time. Since it is not feasible to give effort to every component in the system, there will be some areas of research that progress at a faster pace. Therefore, it is important to know in the design phase how each of the parts fit to the larger whole, even if some areas of research are mostly place holders for the current development cycle.

4.1 PROBLEM-SOLVING ENVIRONMENTS

Problem-solving environments (PSE) provide capabilities designed to manage the numerous complexities associated with multi-discipline modeling (Gallopoulous et al. 1994, Oliveira et al. 1997). Key functions envisioned for a PSE include: 1) constructing the domain and problem specific tools from components, 2) enabling the incorporation of models/components for integrated use with other components, and 3) facilitating the use of visualization, analysis, GIS, and decision support tools, thereby developing an expert system.

4.2 DATA REPRESENTATION

An important research area critical to the success of PSEs is adequate representation of the data to support scale and geographically-dependent scientific process models. Several mathematical concepts form the basis for development of abstractions for handling several prevalent scientific data types. These are fiber bundles (Butler and Pendley 1989, Butler and Bryson 1992) for field data, lattices (Scott 1984) for topology, and cell complexes (Ambrosiano et al. 1999) for meshes. These data models and associated Application Programming Interface (API) libraries can provide a means for chemical, physical, and biological process models of multiple, disparate temporal and spatial specification to intelligently exchange data. The concept of common, shared, data input/output libraries implemented using abstract data types with self-describing file headers offers the potential to overcome many of today's difficulties with data exchange among models from different scientific disciplines.

4.3 CROSS-PLATFORM COMPUTING AND DATA MANAGEMENT

Design of multi-discipline applications from compatible components is based on having a flexible framework that is accessible to a large number of clients. Global shared information and access is achieved via the largest WAN, the Internet. Taking advantage of this network, the framework design allows for database management and use of models and data in a distributed fashion, no matter where they may physically reside. Though originally designed on UNIX workstations and scalable architectures, cross-platform computing and a consistent client interface have been demonstrated on personal computers running Microsoft™ NT and the Models-3 client software. This accomplishment improves the network capabilities of the framework, including data management and access via different platforms. Taking this even further, significant progress has been made in using networks of both PCs and workstations to function in parallel for certain intensive computing applications.

4.4 TECHNOLOGY TO PROMOTE INTERDISCIPLINARY MODULE AND FRAMEWORK EVOLUTION

What we have outlined is a technology whose success depends on an efficient method for community involvement and software evolution. Ultimately, the framework will perform three critical functions: 1) locate a set of candidate components relevant to a specific problem, 2) evaluate the set of retrieved components to find the most suitable, and 3) adapt the selected components when necessary to fit specific requirements. It should be no surprise that these capabilities are identical to the needs of any research and development program following a reuse development cycle for software reuse (Hallsteinsen and Paci 1997).

There is also a similar need for an adaptive, cyclic process of software development in this multimedia research. Much of what is to be attempted with this component architectural framework and scientific module development effort will be tested in a process of adaptively scoping, refining, and implementing the design in several iterations. We do not intend the first demonstration to be the final design, as much of the experience and ultimate success can only be achieved by actually taking the first steps and then gauging the progress. There will also be a degree of tension between the desire to maintain the most flexible and maintainable architecture as possible and the need to emphasize the scientific needs of the various modules within MIMS. Given the data requirements of the modules and their inherent design characteristics (space and time scale(s) of interest, time increment(s), boundary conditions, etc.), compromises in the early stages of development will be crucial to the demonstration of such a dynamic software design.

4.5 NEW NUMERICAL SOLUTION TECHNIQUES

The MIMS will provide a platform for testing innovative numerical solution techniques. For example, the analytic element method has emerged as a tool for describing groundwater flow at multiple (continuous or nested) spatial scales within shallow aquifer systems (Strack 1989, Haitjema 1995). The AEM is based on the superposition of many closed-form analytic solutions to represent hydrological features: point-sinks for wells, line-elements for rivers, area elements for piece-wise or continuous recharge/leakage or other aquifer properties. Multi-scale technology is needed to facilitate the exchange between meso-scale atmospheric models, the stream reaches of the surface water models, and the subsurface models. Hybrid techniques, such as the combination of AEM and finite difference methods, will likely play a role in resolving temporal simulations.

4.6 TESTING THE SYSTEM IN THE FIELD

At the current time we are building and testing the "proof of concept" for the physical, chemical, and biological components of the MIMS in the Albemarle-Pamlico basin (North Carolina, Virginia). A multiple spatio-temporal scale approach is being designed, from site scale (Lizzie, NC) to catchment (Contentnea River, Neuse River) to basin (Albemarle-Pamlico) to study the atmospheric-hydrospheric linkage. A step-wise and progressive development approach is anticipated. We are initially focusing on linking air, surface and groundwater and the simulation of a conservative tracer. We will then add the chemical transformations of nutrients and the transport of particles/sediments. Ecological modeling tools will be implemented as they are developed.

4.7 CONCEPTUAL DEVELOPMENT OF A RISK ASSESSMENT FRAMEWORK

Ecological risk assessment is a field in its infancy, focusing to date primarily on toxic hazards and aquatic endpoints. The definition of sustainable, edible fisheries in the Neuse estuary as an assessment endpoint, with greater complexity than a single species or species-by-species approach can yield, is a formidable challenge. This will involve the integration of information gathered at various scales concerning system structure and function. A conceptual framework for performing ecological risk assessments will be implemented as a combination of models, visualization and analysis tools, and a decision support system for making inferences between various datasets and model results. As an example, satellite data of the assessment area may be combined with output of the physical system simulation previously discussed, incorporating results of laboratory and field studies that tested the effect of multiple stressors on several levels of system organization.

The extension of the scope of this risk assessment to include non-toxic chemical stresses, such as land use change and nitrogen enrichment, is requisite to managing the estuary and the freshwater systems in the Neuse River Basin given the significance of how these disturbances alter hydrologic balances, habitat characteristics, and the structure of ecological communities. The separation of intrinsic variability in the status of aquatic communities from those variations that result from anthropogenic sources of disturbance is also not trivial. It will need to be determined if there are specific classes of problems that can be handled in a similar fashion, as well as how many endpoints must be defined to accomplish an assessment of the entire river basin, including the various lakes, stream, and pond systems. Management alternatives would also be evaluated within the problem-solving environment based on the costs of remediation and other activities and the projected changes in resource quality. Eventually, terrestrial endpoints will be added as well.

One of the goals of a MIMS is such that many functional models can be constructed, from simple to complex, to test the effect of aggregation of model compartments and the degree and type of connectivity required to reproduce cause-and-effect relationships, given data limitations and scientific uncertainties in performing risk assessments. Parameter sensitivity analysis is one of the most common ways in which individual modules are tested; a major challenge for MIMS will be evaluation of entire systems of modules.

There will always exist a need for a synthesis of science and technology for performing risk assessments, and it seems this is best achieved in an evolutionary (e.g., object oriented) software design using problem-solving environments and facilitated visualization of model results in comparison to field observations.

5. Conclusions and Beginnings

What we have described will hopefully encourage others to become involved in the conceptual model building exercise, as well as the long term software development. Problem-solving environments are both a helpful concept for organizing the manner in which one approaches complex system modeling and a useful consequence of object-oriented software development, with obvious benefits for model building. A community effort from researchers is clearly needed to effectively tackle the challenge of complex system modeling and ecological problem-solving. With its cross-platform compatibility and accessibility via the Internet, a common framework holds the promise of being an inclusive research environment and a means of community collaboration. The EPA's goal of a comprehensive, multimedia modeling system is an attempt to deal with complex environmental problems in a concerted and realistic fashion, while realizing that the task is greater than the capabilities of any one organization.

Acknowledgments

The manuscript was improved by the comments of Robert Swank (U.S. EPA, Ecosystem Research Division, Athens, GA) and two anonymous reviewers. This paper has been reviewed in accordance with the U.S. Environmental Protection Agency's peer and administrative review policies and approved for presentation and publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

Ambrosiano, J., Butler, D.M., Matarazzo, C., Miller, M. and Schoof, L.: 1999, *Development of a common data model for scientific simulations,* Los Alamos Technical Report LA-UR-99-722, submitted to the 11th International Conference on Scientific and Statistical Database Management, Cleveland, OH, July, Los Alamos National, July 1999.

Bachman, L.J., Lindsey, B., Brakebill, J. and Powars, D.S. : 1998, *'Ground-water discharge and base flow nitrate loads to non tidal streams and their relations to a hydrogeomorphic classification of the Chesapeake Bay watershed'*, middle Atlantic Coast, U.S. Geological Survey, Water-Resources Investigations Report 98-4059.

Butler, D.M. and Pendley, M.H.: 1989, 'A Visualization Model Based on the Mathematics of Fiber Bundles', *Computers in Physics* **3**(5), September/October 1989.

Butler, D.M. and Bryson, S.: 1992, 'Vector-Bundle Classes Form Powerful Tool for Scientific Visualization', *Computers in Physics* **6**(6), 576–584.

Gallopoulous, E., Houstis, E. and Rice, J.R.: 1994, 'Problem-Solving Environments for Computational Science', *IEEE Computational Science and Engineering,* Summer 1994, pp. 11–23.

Haitjema, H.M.: 1995, *Analytic Element Modeling of Groundwater Flow*, Academic Press, San Diego, California.

Hallsteinsen, S. and Paci, M. (eds.): 1997, *Experiences in Software Evolution and Reuse*. Springer, Berlin.

Oliveira, J.L., Pires, F. and Medeiros, C.B.: 1997, 'An Environment for Modeling and Design of Geographic Applications', *GeoInformatica,* **1**, 29–58.

Scott, D.S.: 1984, 'Data Types as Lattices', *SIAM Journal of Computing* **5**(3), 522–587.

Strack, O.D.L.: 1989, *Groundwater Mechanics*, Prentice Hall, Englewood Cliffs, New Jersey.

Theodore, M.K. and Theodore, L.: 1996, *Major Environmental Issues Facing the 21st Century,* Prentice Hall, Englewood Cliffs, New Jersey.

USEPA: 1998, *Third-generation Air Quality Modeling System, Models-3 Volume 9b: User Manual, EPA/600/R-98/069(b),* U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC, 833 pp.

USEPA: 1999, *Science algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) modeling system,* Byun, D.W. and Ching, J.K.S. (eds.), EPA-600/R-99/030, U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC.